# ADAPTIVE WEB CRAWLING USING A STATISTICAL MODEL


## Abstract of the Disclosure

A computer based system and method of retrieving information pertaining to documents on a computer network is disclosed. The method includes selecting a set of documents to be accessed during a Web crawl by utilizing a statistical model to determine which previously retrieved documents are most likely to have changed since last accessed. The statistical model is continuously improving its accuracy by training internal probability distributions to reflect the actual experience with change rate patterns of the documents accessed. The decision made whether to access the document is based on the probability of change compared against a desired synchronization level, random selections, maximum limits on the amount of time since the document was last accessed, and other criterion. Once the decision to access is made, the document is checked for changes and this information is used to train the statistical model.